2009 AVSS Multiple Camera Person Tracking (MCPT) Evaluation Plan

1 Introduction

Technologies that extract information from video sensors are being used for a variety of security applications from a variety of domains such as mass transit monitoring, large entertainment venue security, building security, etc. A key component technology for these domains is the ability to track people as they move through a network of video cameras. Several previous evaluations have focused on different aspects of person tracking challenges including: Classification of Events, Activities, and Relations (CLEAR) [3], Performance Evaluation of Tracking and Surveillance (PETS) [4], TRECVid's Event Detection Track [6].

To continue the effort to develop person tracking technologies, the Advanced Video and Signal Based Surveillance (AVSS) IEEE Conference is sponsoring the Multiple Camera Person Tracking challenge evaluation in conjunction with the Home Office Scientific Development Branch (HOSDB), Centre for the Protection of National Infrastructure (CPNI), and the National Institute of Standards and Technology (NIST).

The goal of the effort is to facilitate research via a common evaluation task that focuses on one aspect of person tracking technologies: the ability to track a specified person within a video sensor field using a small set of in situ exemplar video images to specify the person. We refer to these technologies as **Single Person Tracking (SPT)** technologies.

The rest of this document describes the evaluation tasks supported by the evaluation, the data set provided to the participants at no charge, the metrics used to evaluate system performance, data formats, and system submission instructions.

2 The Evaluation Tasks

The evaluation supports three evaluation tasks: Multi-Camera Single Person Tracking (MCSPT), Single Camera Single Person Tracking, and Camera Pair Single Person Tracking. The first task, MCSPT, is the compulsory task that all participants must build systems address. The latter two are voluntary, contrastive evaluation tasks designed assess factors of SPT system performance.

2.1 Multi-Camera Single Person Tracking (MCSPT) Task

The Multi-Camera Single Person Tracking task is to spatio-temporally track a single person as they traverse a multi-camera field after the person to track has been specified by five in situ video frames. The in situ video frames, called ***Target Tracking Frame(s) (TTF)***, will be selected from a single camera view and the frames will be the first five annotated frames where the subject is 100% within the frame boundaries (75% for elevator close-up camera) and with no greater than 50% occlusion. The system must then track the person in all video streams from the next frame on.

Implicit in the MCSPT task, systems are expected to be able to re-acquire the subject during camera transitions that may or may not overlap and regardless of whether or not the system losses a track.

2.2 Single-Camera Single Person Tracking (SCSPT) Task

The Single-Camera Single Person Tracking task is to spatio-temporally track a single person as they move through a single camera view after the person to track has been specified by five TTFs. The system must then track the person after the last TTF for the remainder of the video.

This evaluation task is a contrastive condition to determine the system's ability to track the person within camera by factoring out camera-to-camera target re-acquisition of the MCSPT task.

2.3 Camera-Pair Single Person Tracking (CPSPT) Task

The Camera Pair Single Person Tracking task is to spatio-temporally track a single person as they traverse two camera fields of view after the person to track has been specified by five TTFs. The system must then track the person in the pair of video streams.

This evaluation task is a contrastive task focusing on the system's ability to successfully re-acquire the target between a pair of cameras. The "second" camera in the pair may or may not contain images of the person and the camera's field of view may or may not overlap.

3 The Data

The training and test data for the evaluation comes from HOSDB's i-LIDS Multiple Camera Tracking Training Corpus (MCTTR) [1]. The data set was collected at the London Gatwick Airport from a 5-camera airport surveillance field and it has a total of ~44 camera hours of data. The collection consists 107 5-camera excerpt sets from 12 collection epochs. Each excerpt

contains a track for a single person annotated according to the guidelines in the i-LIDS User Guide [2].

The data set will be divided into two subsets for the AVSS MCPT evaluation: a ~29-hour training corpus and a ~15 testing corpus. The data will be divided by using the following prioritized factors:

- Excerpts from a collection epoch will not be split across test/train data sets,
- Subject trajectories through the camera network will be represented in both test/train data sets
- Crowd density will be balanced
- Time of day will be balanced
- Gender of tracked people will be balanced

The MCTTR data set has already been exposed in that some participants may have worked with the data already. As such, participants will be expected to state the extent of their use of the video evaluation data prior to developing a system for AVSS and any system output submitted for AVSS may not make use of the test data for system training/tuning.

4 The Metrics

Systems will be evaluated with several different metrics. The primary evaluation metric is TBD but expected to be one of following:

- F1 (as described in HOSDB Pub 28-08) [2]
- Multiple Object Tracking Accuracy [3, 5]
- Multiple Object Tracking Precision [3, 5]
- Normalized Detection Cost Rate [6]

This section will further describe the protocols and metrics at a later date.

5 Data Formats

The reference annotations and system outputs will be ViPER formatted XML files following the conventions used for the CLEAR Person Tracking evaluation. Appendix A (which will be added later) will document the data format.

6 System Output Submission Instructions

This section will be completed at a later date.

7 References

1. Home Office Multiple Camera Tracking Scenario data, http://scienceandresearch.homeoffice.gov.uk/hosdb/cctv-imaging-technology/video-based-detection-systems/i-lids
2. i-LIDS User Guide: http://scienceandresearch.homeoffice.gov.uk/hosdb/publications/cctv-publications/28-08_-_i-LIDS_User_Guide.pdf
3. 2007 CLEAR Evaluation Protocol, http://isl.ira.uka.de/clear07/downloads/ClearEval_Protocol_v5.pdf
4. PETS: Performance Evaluation of Tracking and Surveillance, http://www.cvg.rdg.ac.uk/slides/pets.html
5. Kasturi, Goildgof, Soundararajan, Manohar, Garofolo, Boonstra, Korzhova, Zhang, "Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 2, Feb. 2009
6. TRECVid Event Detection Evaluation Plan: http://www.nist.gov/speech/tests/trecvid/2008/doc/EventDet08-EvalPlan-v07.htm